

# ADVANTAGES AND DISADVANTAGES OF DIMENSIONALITY REDUCTION METHOD

## ① Principal Component Analysis -

- a) Advantages -
- i) Remove correlated features.
  - ii) Reduce overfitting.
  - iii) Improving algorithm performance (With so many features, performance decrease drastically in which PCA helps)
  - iv) Improve Visualization (PCA transforms a high to low dimensional 2D which we can visualize easily)
  - v) Overfitting mainly occurs when there are too many variables. So PCA helps by reducing number of features.

- b) Disadvantages -
- i) Independent variable become less interpretable. Original feature turn into Principal Component. Principal component are combination of original feature. So PC are not as readable and interpretable as original feature.

ii) Data standardization is must before PCA

iii) Information loss - Although PC try to cover maximum variance, we may miss some information.

## ② T-SNE ->

Advantages - ① Though PCA is great, it does have some severe drawbacks. One drawback of PCA is that it is a linear projection, meaning it can't capture non-linear dependencies. Unlike PCA, t-SNE is not limited to linear projections, which makes it suited to all sort of the datasets.

② Major benefit of t-SNE is that it uses "stochastic neighbors". This means there are no clear line between which points are neighbors of the other point. This lack of clear borders can be a major advantages because it allows t-SNE to naturally take both global and local structure into account.

Disadvantages - ① Non convexity of optimization, unlike methods like PCA, t-SNE is non-convex meaning it has multiple local minima and is therefore much more difficult to optimize. T-SNE is non-deterministic we can run it multiple times and get a different result.



② Complex manifolds, though t-SNE is not a linear projection, it still does not make some important assumption about the data. One assumption is that the local structure of the manifold is linear. The reason this assumption is ~~linear~~ important is that distance between neighboring points is measured in Euclidean distance which assume linearity. When local structure is complex, this poses a significant problem.

### ③ Factor Analysis -

Limitations - i) Random data give factors, if we generate a lot of random numbers, a factor analysis may still find apparent structure in data. It's difficult to tell if the factor emerge reflect the data or are simply part of power of factor analysis.

- ii) It is hard to decide how many factors to include, there are variety of methods but there are little agreement as to which is best.
- iii) Interpretation of meaning of factor is subjective, factor analysis can tell which variable in dataset go together in way they aren't obvious. But interpreting what those set of variable actually represent may be subjectable.

### ④ LDA -

Advantages of LDA over logistic -

- 1) Two class problem, Logistic regression used for binary classification. LDA is considered a better choice for multi-class classification.
- 2) Unstable with well separated class, logistic regression lack stability when the classes are well separated. This is where LDA comes in.
- 3) If there are few examples from which parameters are to be estimated, logistic regression become unstable. However LDA is better option because it tends to be stable even in such cases.

PCA	Factor Analysis	PCA	LDA
① PC component explain maximum amount of variance.	① FA explain maximum amount of covariance.	① Ignore class label, maximize variance.	① Based on class label, maximize separation between class labels.
② Kind of dimensionality reduction technique.	② Kind of latent variable method.	② Perform well than LDA when sample size small	② LDA perform better multi-class classification than PCA.
③ Component are uninterpretable	③ Factors are labeled & interpret	③ Unsupervised algorithm	③ Supervised algorithm
④ Linear combination of observed variable.	④ Linear combination of Unobserved variable / factor.		
⑤ Component are fully orthogonal to each other	⑤ Does not require factors to be fully orthogonal.		